" AN ALTERNATIVE APPROACH TO THE QUESTION OF IONIC BALANCES "

WESTERN & NORTHERN REGION

WATER QUALITY BRANCH

JAMES E. GASKIN

TABLE OF CONTENTS

		PAGE
1.	INTRODUCTION	1
2.	CONTROL CHARTS	2
3.	THEORY	3
4.	ALTERNATIVE APPROACH TO IONIC BALANCES	6
5.	SUMMARY	12
	REFERENCES	14

1. INTRODUCTION

Theoretically, the sum of the anions, expressed in milliequivalents per litre (me/L), must equal exactly the sum of the cations, similarly expressed, in any given sample. In practice, the sums seldom are equal because of unavoidable variations in analysis. This inequality increases as the ionic concentration increases.

Another possible source of error in the ionic balance equation may arise because only the so called major ions* (Na, K, Ca, Mg, Cl, F, SO_4, NO_3, NO_2 and SiO_2) are considered, whereas ions such as Fe, Cu, PO_4, etc. which are relatively ubiquitous, are omitted from the balance calculations. The size of this error is hard to predict, but it might be of importance in some individual cases, depending on the source(s) from which the water samples are taken.

2. CONTROL CHARTS

Variation produced by chance causes follows statistical laws. Knowledge of the behaviour of chance variations is the foundation on which Control Chart Analysis rests. If a group of data is studied and it is found that their variation conforms to a statistical pattern that might reasonably be produced by chance causes, then it is assumed that no special assignable causes are present. The conditions which produced this variation are, accordingly, said to be ' under control'. They are under control in the sense that, if chance causes are alone at work, then the amount and character of the vari-

* CO_3 and HCO_3 ions are also included.

.../2

ation may be predicted for large numbers, and it is not possible to trace the variation of a specific instance to a particular cause. On the other hand, if the variations in the data do not conform to a pattern that might reasonably be produced by chance causes, then it is concluded that one or more assignable causes are at work. In this case, the conditions producing the variation are said to be 'out of control'.

A control chart is a device for describing in concrete terms what a state of statistical control is ; secondly, it is a device for attaining control; and thirdly, a control chart is a device for judging whether control has been attained.

In the Standard Methods Manual[1], a control chart was constructed to indicate if the difference between the sums of the anions and cations fall between acceptable limits, which have been taken as ±1 standard deviation.
When the difference (Σ[Anions] - Σ[Cations]) was plotted against the sum of the anions (Σ[Anions]), lines showing ±1 standard deviation (considered as acceptable limits), were given by the equation :

$$\Sigma[\text{Anions}] - \Sigma[\text{Cations}] = \pm (0.1065 + 0.0155\Sigma[\text{Anions}]) \quad \text{--------} \quad (1)$$

This equation has become the standard equation in assessing the nature of the ionic balances in water samples from the Western and Northern Region of the Water Quality Branch, despite the existence of some difficulty in determining the exact derivation of the equation , and its relative applicability to the analytical results for the major ions in the W&NR.

In this exercise we have thus set about to outline the development of the

theory that supports the concept of plotting a control chart for analytically measured values. In addition, we have taken a slightly different approach in effecting our correlation between ($\Sigma A - \Sigma C$) and ΣA.

3. THEORY

The regression line of y on x can be written in the general form as :

$y = a + bx$, where a is the intercept and b is the slope of the line.

Values of y can be estimated from the regression line for each corresponding given value of x; i.e., $y_E = a + bx_G$ ---------- (2).

The population composite variance, which is sometimes called the Standard Error of Estimate (SEE), is then related to both the observed and estimated y values, through the following equation :

$$\sigma^2_{y,x} = \frac{\Sigma(y_0 - y_E)^2}{N} \qquad \text{---------- (3)}$$

where N is the number of data points.

Also $(y_{obs.} - y_{est.}) = y_{obs.} - (a + bx_{given})$ ---------- (4)

Analysis of equation (3) allows us to re-write $\sigma^2_{y,x}$ in the form,

$$\sigma^2_{y,x} = \frac{\Sigma y^2 - a\Sigma y - b\Sigma xy}{N} \qquad \text{---------- (5)}$$

The use of either equation (3) or equation (5) can be demonstrated by considering the following example :

The following are measurements of a chemical precipitate (y) and a reagent (x) used in obtaining this precipitate.

TABLE 1

x	y	x	y	x	y
8.2	9.4	7.1	9.3	8.6	14.2
5.8	6.5	7.8	10.6	9.0	13.8
6.4	7.3	8.1	11.4	8.1	11.3
5.9	7.9	7.5	12.2	5.7	8.1
6.5	9.0	7.9	13.3	5.3	6.1

Compute the standard error of estimate (composite error) $\sigma^2_{y,x}$ for the data.

SOLUTION :

To use either formula (3) or (5), the regression line of y on x must first be determined. The computed equation to the regression line was found to be,

$$y = -4.317 + 1.994x$$

Having obtained the regression line, the values of the a and b coefficients can be used in formula (5), or we could compute $y_{est.}$ (estimated values of y) from the regression line for each corresponding x value, and then insert

these values in equation (3). The second approach has an added bonus as we shall see later.

Let us construct a Table for $y_{est.}$ and ($y_{obs.} - y_{est.}$) values.

TABLE 2

Comparison of $y_{obs.}$ values with $y_{est.}$ values

x	8.2	5.8	6.4	5.9	6.5	7.1	7.8	8.1	7.5	7.9	8.6	9.0	8.1	5.7	5.3
$y_{obs.}$	9.4	6.5	7.3	7.9	9.0	9.3	10.6	11.4	12.2	13.3	14.2	13.8	11.3	8.1	6.1
$y_{est.}$	12.0	7.2	8.4	7.4	8.6	9.8	11.2	11.8	10.7	11.5	12.8	13.6	11.8	7.0	6.3
$y_o - y_e$	-2.6	0.7	1.1	0.5	0.4	-0.5	-0.6	-0.4	1.5	1.8	1.4	0.2	-0.5	1.1	-0.2

Example : $y_{est.} = -4.317 + 1.994 (8.2)$ for the first x- value (8.2)

i.e. $y_{est.} = 12.0$

$$\sigma^2_{y,x} = \frac{\Sigma(y_{obs.} - y_{est.})^2}{N} = 3.033 ; \quad \sigma_{y,x} = 1.74 .$$

If we were dealing with a sample of the readings, an estimate of the standard error would have been given by

$$s^2_{y,x} = \left(\frac{N}{N-2}\right) \sigma^2_{y,x} \qquad \text{----------} \qquad (6)$$

where N-2 is the actual number of degrees of freedom .

.../6

The next step in the evolution of the theory, is to construct two lines parallel to the regression line, one on each side , and each displaced a vertical distance of one standard deviation from the estimated (regression) line. The three lines are shown in Figure 1.

From Figure 1. it can be seen that of the 15 data points, 13 fall between the lines, while 2 fall beyond the boundaries of the outer lines. The percentage of points falling between the lines is [(13/15)x100] = 86.7% .

From the last line in Table 2, there are 13 points that lie between -1.74 and +1.74 (i.e. $\pm \sigma_{y,x}$), also producing an 86.7% coverage.

If the points were normally distributed about the regression line, theory predicts that about 68% of the points would lie between the lines. This would have been more nearly the case if the population or sample size were larger; possibly in the vicinity of 30 observations(ν = 28 degrees of freedom* for a sample size of 30).

4. ALTERNATIVE APPROACH TO IONIC BALANCES

At this stage we were still labouring under an emerging conviction that the formula in the Standard Methods Manual for judging the quality of ionic balances, was not the most appropriate vehicle for our calculations in the W&NR. We therefore needed at least a slightly modified approach.

* The 'number of degrees of freedom' of a statistic generally denoted by F or ν is defined as the number of N independent observations in the sample size minus the number k of population parameters which must be estimated from sample observations. In symbols, $\nu = N-k$.
In the above case, $\bar{x}$ and $\bar{y}$ are the two estimated parameters; these two parameters constitute the centroid, through which the regression line must pass.

.../7

Presently, ionic balances judged by using the Standard Methods Formula, are labelled either 'good', 'poor' or 'bad'. It is apparent that a number of uncontrollable and undetermined factors could collectively be the cause of a so called poor or bad balance. We have already mentioned factors such as the absence of the consideration of the contribution of other ions, namely, Fe, PO_4, etc.

There are times when ionic balances may also be fortuitously good due to a combination of erroneous analyses resulting in the balancing of errors. The Manual suggests additional methods of checking inorder to uncover any discrepancies that might have occurred in the analyses[2].

The use of $\pm 1\sigma$ as control limits for ionic balances in the light of probable attendant errors during analyses and the possible omission of other important contributing factors, does appear somewhat stringent and tight. Perhaps $\pm 1\sigma$ could be used as 'inner' or 'warning' limits and $\pm 2\sigma$ could be used as ordinary ' action ' limits. Where points fall preponderantly outside the action limits, tighter experimental control and additional data would be advised.

To test the model that we have been proposing, we decided to look at a randomly selected sample of ionic balances which were computed for water samples gathered in the Western and Northern Region over the past three years. From hundreds of ionic balances, 155 ionic balances were chosen at random. From the group of these mixed-labelled ionic balances, 30 ionic balances were randomly sampled, using computer generated Tables of random units. In statistical reasoning, a sample size of 30 or more items is considered

.../7a

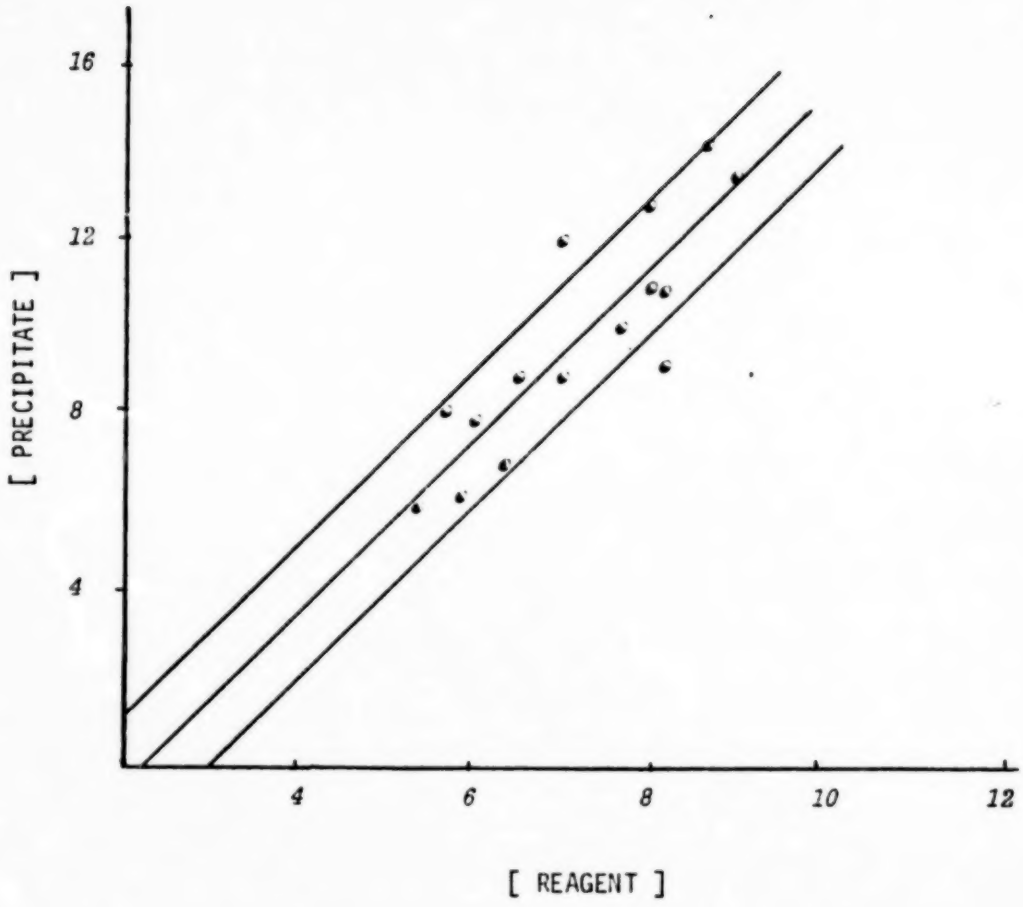

Figure 1. A plot of concentration of precipitate versus concentration of reagent.

.../8

large enough for a sampling distribution to be reasonably normal (Bell-shaped).

Over the past three or more years, the ionic balances (W&NR) labelled ' good ' has been much larger than those proportions labelled ' poor ' and ' bad '. This was reflected in the proportions obtained for each category in the 155 randomly chosen samples, and then in the final 30 randomly selected samples (Table 3).

TABLE 3

Display of Proportions by Categories of Ionic Balances in two randomly chosen sets

Category	Proportions from 155 Samples	Proportions from 30 Samples
GOOD	0.50	0.50
POOR	0.28	0.30
BAD	0.22	0.20

When the 155 samples were scrutinized, it was noticed that some ionic balances which were labelled 'good', had a ratio of $\frac{\Sigma A}{\Sigma C}$ or $\frac{\Sigma C}{\Sigma A}$ of approximately 0.827, and some labelled 'poor' or 'bad' had ratios in the region of 0.970 (Table 4). In addition some of the ($\Sigma A - \Sigma C$) values labelled 'good' had the same range as those labelled 'poor' and 'bad'.

There was no explanation in the references[3,4] as to what criteria were used to separate similar differences. Ratios were not considered or referenced in

.../9

the Standard Methods Manual.

TABLE 4

Cationic and Anionic Differences and Ratios for Ionic Balances for 30 randomly chosen samples (WQB, W&NR).

$(\Sigma C-\Sigma A)$	$\frac{\Sigma A}{\Sigma C}$ or $\frac{\Sigma C}{\Sigma A}$	$(\Sigma C-\Sigma A)$	$\frac{\Sigma A}{\Sigma C}$ or $\frac{\Sigma C}{\Sigma A}$
-0.014	0.991 (G)	0.342	0.963 (P)
0.090	0.961 (G)	-0.080	0.968 (G)
0.333	0.938 (P)	-0.141	0.921 (P)
0.076	0.949 (G)	0.228	0.931 (P)
-0.390	0.969 (P)	0.256	0.913 (P)
0.216	0.956 (P)	-0.857	0.907 (B)
0.130	0.925 (G)	0.432	0.915 (B)
0.111	0.951 (G)	0.333	0.946 (P)
0.024	0.991 (G)	0.215	0.950 (P)
0.324	0.975 (G)	0.070	0.972 (G)
0.595	0.940 (B)	0.759	0.909 (B)
0.043	0.868 (G)	0.086	0.975 (G)
0.353	0.898 (B)	0.282	0.967 (G)
0.043	0.972 (G)	-0.707	0.952 (B)
0.006	0.996 (G)	0.050	0.827 (G)

It was felt that by correlating individual ratios and differences, a better estimate for ionic balance categorization could be obtained in terms of one to three standard deviations from the Theoretical Mean Difference Line (c.f. Figure 2) - $[(\Sigma C - \Sigma A)] = 0$.

.../9a

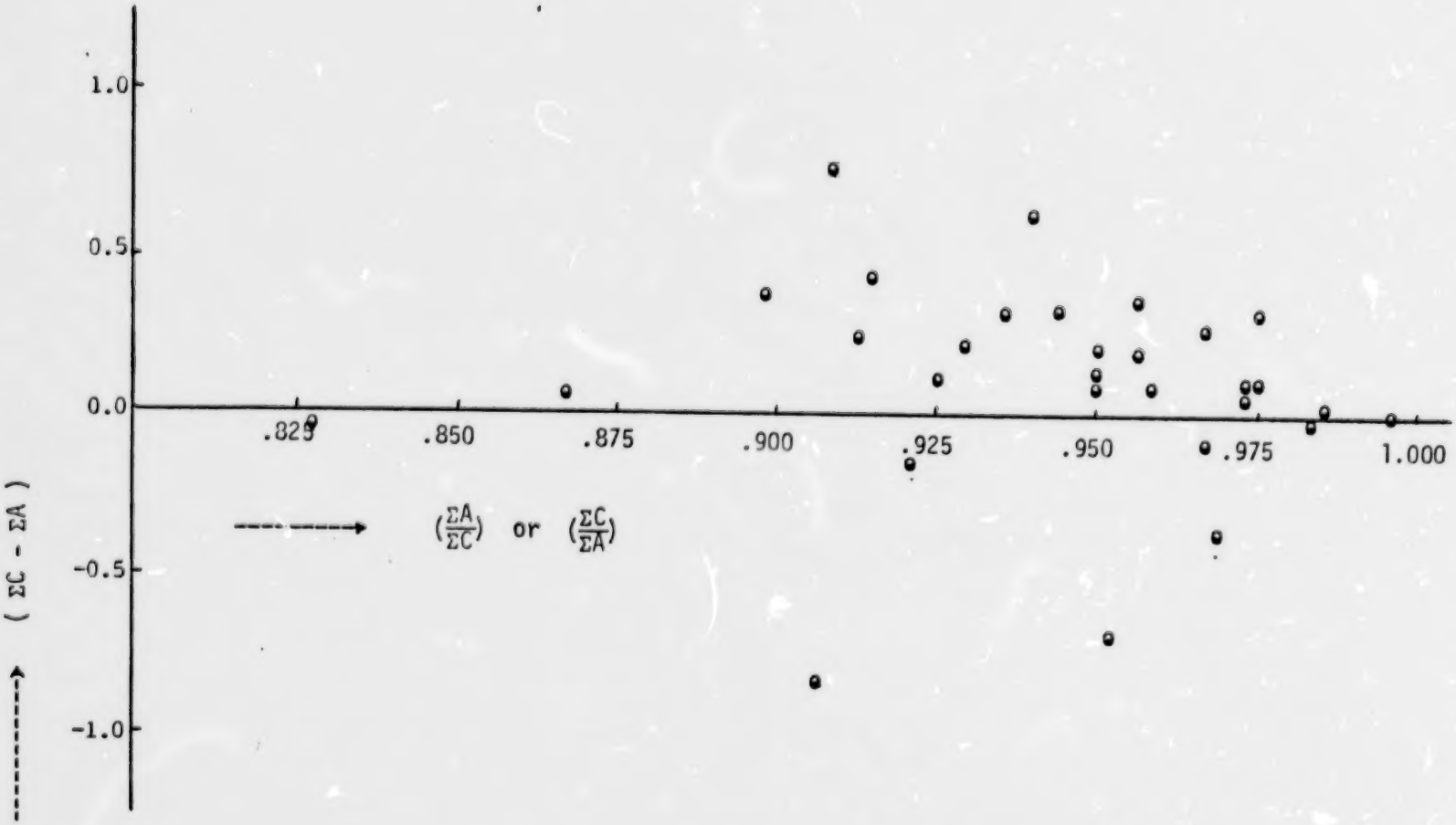

Figure 2. Correlation of (ΣC - ΣA) differences with $\frac{\Sigma C}{\Sigma A}$ or $\frac{\Sigma A}{\Sigma C}$ values

.../10

The calculations conducted on the 30 randomly sampled ionic balances gave the following results

$$a = 0.1047$$

$$b = 0.0026$$

for the estimated line

$$(\Sigma C - \Sigma A)_{est.} = a + b\left(\frac{\Sigma C}{\Sigma A}\right) \text{ -------- (7)}^*$$

$$(\Sigma C - \Sigma A)_{est.} = a + b\left(\frac{\Sigma A}{\Sigma C}\right) \text{ --------- (8)}^*$$

$\Sigma y^2 = 3.4501$; $\Sigma y = 3.2070$; $\Sigma(xy) = 3.0007$

where $y = (\Sigma C - \Sigma A)$ and $x = \frac{\Sigma A}{\Sigma C}$ or $\frac{\Sigma C}{\Sigma A}$

$$\sigma^2_{y,x} = \frac{\Sigma y^2 - a.\Sigma y - b\Sigma(xy)}{N}$$

$$= \frac{3.4501 - (0.1047)(3.207) - (0.0026)(3.0007)}{30} = 0.1041$$

* Equations (7) and (8) are interchangeable depending on whether $\frac{\Sigma A}{\Sigma C} < |1|$ or $\frac{\Sigma C}{\Sigma A} < |1|$.

.../11

$$\sigma_s^2 = s_{y,x}^2 = \left(\frac{N}{N-2}\right)(0.1041) = \frac{30}{28}(0.1041) = 0.1115$$

$$\therefore \quad \sigma_s = 0.334$$

Figure 2 shows that 8 points for the ($\Sigma C - \Sigma A$) differences lie beyond the $\pm 1\sigma_s$ limits, whereas two points lie on the line, and 20 points lie within the boundaries of the Mean Difference Line (MDL) and the $\pm 1\sigma_s$ limit lines. The population of 22 points translates into a percentage of 73.3%. A similar percentage is obtained when the ($\Sigma C - \Sigma A$) differences in Table 4 are compared to $\pm 1\sigma_s$.

If the points were normally distributed about the Mean Difference Line, then 68% of the points should theoretically lie between the lines. If we increase the final sample size from 30 to say 100 ionic balances, a situation resembling a normal distribution of points would be more attainable.

A plot of the regression line in Figure 3 shows the line to be 'a' units (0.105) removed from the MDL. If vertical distances were measured in the positive direction from the regression line which is approximately parallel to the MDL, points which are o.439 units (0.334 + 0.105) away from the MDL could be accommodated in the vicinity of $\pm(\sigma_s \pm a)$ for a 'good' or acceptable ionic balance, if a 'good' limit is accepted as $(\Sigma C - \Sigma A) \leqslant 1.5\sigma_s$.

It is therefore proposed that the coding in Table 5 be adopted for future assessments of Ionic Balances derived from major ions, nutrients and alkalinity analyses for water samples from the Western and Northern Region.

.../11a

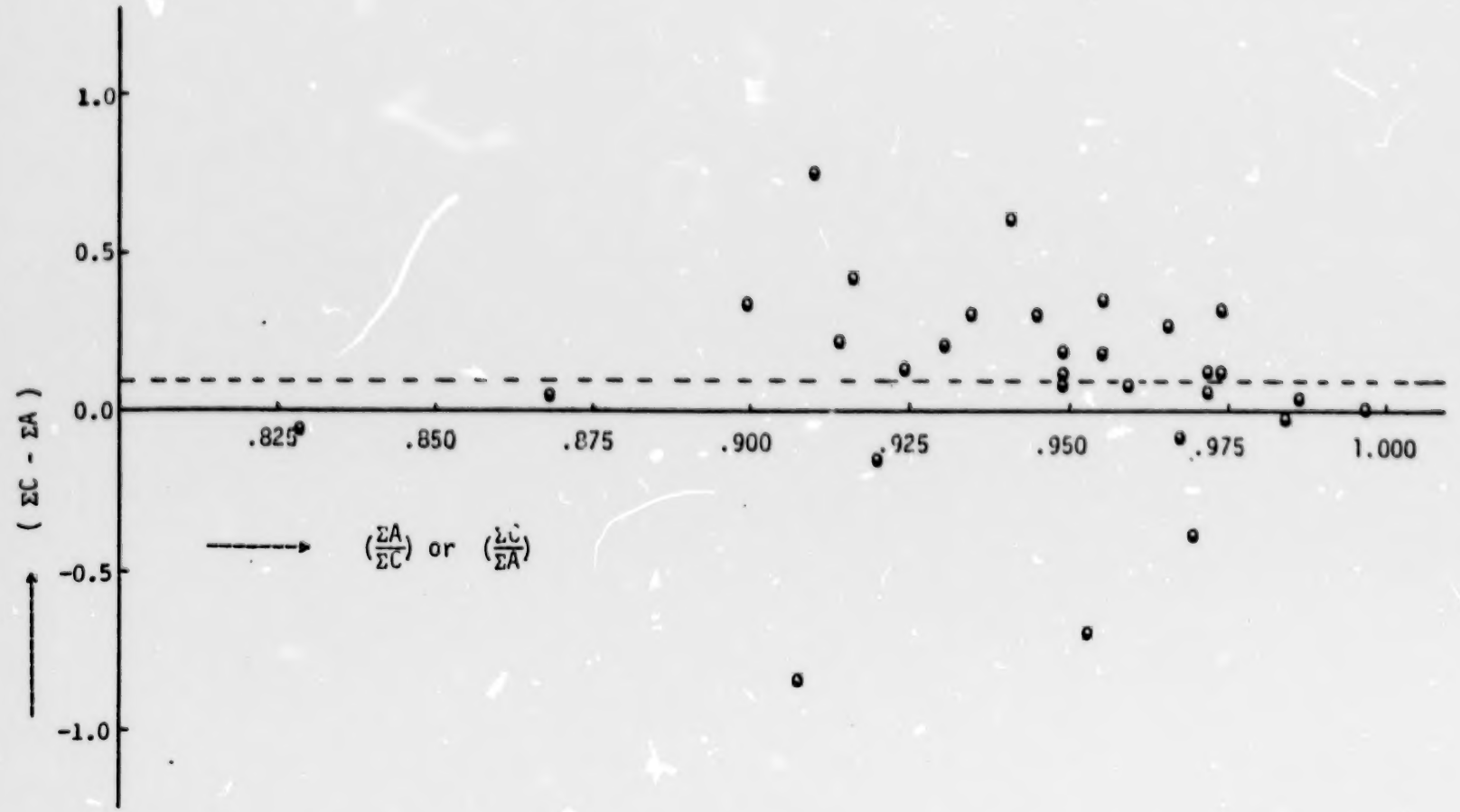

Figure 3. The Regression Line in relationship to the Mean Difference Line.

.../12

TABLE 5

Coding for Ionic Balances (W&NR)

#	$\lvert(\Sigma C - \Sigma A)\rvert$	CODE
(a)	$\lvert(\Sigma C - \Sigma A)\rvert \leqslant 1.5\sigma_s$	GOOD
(b)	$1.5\sigma_s < \lvert(\Sigma C - \Sigma A)\rvert \leqslant 2.0\sigma_s$	ACCEPTABLE*
(c)	$\lvert(\Sigma C - \Sigma A)\rvert > 2.0\sigma_s$	POOR

5. SUMMARY

In very serious cases where $\lvert(\Sigma C - \Sigma A)\rvert \geq 3\sigma_s$, it may be necessary to repeat the alkalinity analyses, and possibly some of the measurements of other contributing ions (Na^+, Cl^-, etc.). It is therefore very important that analyses involving alkalinity, major ions, and nutrients measurements be conducted as soon as possible after the collection of samples is made (or delivered to the respective laboratory). In addition, wherever and whenever it is possible, the samples should be well preserved for a suitable and reasonable time period (within the constraints of bottle needs), in case further analytical scrutiny becomes necessary.

The wealth of accumulated data on ionic balances in the records of the

* " ACCEPTABLE " fills the GAP.

.../13

Water Quality Laboratory in the Western and Northern Region, made this a very opportune time for an attempt at deriving a more appropriate assessment of the nature of ionic balances for water samples from the Western and Northern Region, WQB.

.../14

REFERENCES

(1) " Standard Methods, for the Examination of Water and Wastewater ", p.30, 15th Edition, APHA - AWWA - WPCF, 1980.

(2) " Standard Methods, for the Examination of Water and Wastewater ", p.31 - 35, 15th Edition, APHA - AWWA - WPCF, 1980.

(3) McFarren, E.F., Lishka, R.J., and Parker, J.H., " Criterion for judging, and acceptability of analytical methods ", Analytical Chemistry, 42 : 358, 1970.

(4) " Evaluation of Laboratory Methods for the Analysis of Inorganics in water, Advan. Chem. Ser. No. 73, p.253, American Chemical Soc., 1968.